GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:50 ; Search time 91.75 Seconds
(without alignments)
16.543 Million cell updates/sec

Title: US-09-331-631A-7_COPY_34_80
Perfect score: 258
Sequence: 1 YERDPRQQYEQCQRRCESEA.....QCEQRCEREYKEQQRQQEEE 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	124	48.1	588	1	VCLB_GOSHI	P09801 gossypium h
2	124	48.1	605	1	VCLA_GOSHI	P09799 gossypium h
3	87	33.7	429	1	APA4_MACFA	P33621 macaca fasc
4	86	33.3	47	1	AGRP_LUFCY	P56568 luffa cylin
5	86	33.3	1898	1	TRHY_HUMAN	Q07283 homo sapien
6	84.5	32.8	1407	1	TRHY_RABIT	P37709 oryctolagus
7	79.5	30.8	401	1	APA4_PAPAN	Q28758 papio anubi
8	79	30.6	743	1	ABRA_PLAFC	P22620 plasmodium
9	79	30.6	1905	1	TAGB_DICDI	P54683 dictyosteli
10	78	30.2	892	1	IF2_SALTY	Q9zf31 salmonella
11	78	30.2	905	1	SNF5_YEAST	P18480 saccharomyc
12	78	30.2	1403	1	PRO_DROME	P29617 drosophila
13	77.5	30.0	502	1	SKS1_YEAST	Q12505 saccharomyc
14	77	29.8	339	1	TF2D_HUMAN	P20226 homo sapien
15	77	29.8	857	1	NFM_CHICK	P16053 gallus gall
16	76	29.5	482	1	U2R2_HUMAN	Q15696 homo sapien
17	76	29.5	648	1	KAPC_DICDI	P34099 dictyosteli
18	76	29.5	890	1	IF2_ECOLI	P02995 escherichia
19	75.5	29.3	1023	1	GLT_DROME	P33438 drosophila
20	75	29.1	241	1	YO87_CAEEL	P34623 caenorhabdi
21	75	29.1	758	1	YM38_YEAST	Q03825 saccharomyc
22	74.5	28.9	1549	1	TRHY_SHEEP	P22793 ovis aries
23	74	28.7	877	1	INCE_CHICK	P53352 gallus gall
24	73.5	28.5	1344	1	IF3A_MOUSE	P23116 mus musculu
25	73.5	28.5	1382	1	IF3A_HUMAN	Q14152 homo sapien
26	73	28.3	418	1	YSE2_CAEEL	Q09936 caenorhabdi
27	73	28.3	678	1	GARP_PLAFF	P13816 plasmodium
28	73	28.3	1023	1	CLOC_DROME	O61735 drosophila
29	72.5	28.1	287	1	TRT2_HUMAN	P45379 homo sapien
30	72	27.9	407	1	IE68_HSVSA	Q01042 herpesvirus
31	71.5	27.7	462	1	U2R2_MOUSE	Q62377 mus musculu
32	71.5	27.7	1403	1	YDF3_SCHPO	Q10475 schizosacch
33	71	27.5	510	1	CF2_DROME	P20385 drosophila
34	71	27.5	514	1	CF23_DROME	Q01522 drosophila
35	71	27.5	966	1	SSN6_YEAST	P14922 saccharomyc
36	71	27.5	2124	1	Y192_HUMAN	Q93074 homo sapien
37	70.5	27.3	255	1	LP61_EIMTE	P15714 eimeria ten
38	70.5	27.3	503	1	LEGJ_PEA	P05692 pisum sativ
39	70.5	27.3	1391	1	MST2_DROHY	Q08696 drosophila
40	70	27.1	917	1	IF2_PROVU	Q9zf22 proteus vul
41	69.5	26.9	298	1	TRT2_RAT	P50753 rattus norv
42	69.5	26.9	300	1	TRT2_MOUSE	P50752 mus musculu
43	69.5	26.9	466	1	CYP8_CAEEL	P52016 caenorhabdi
44	69.5	26.9	1240	1	YNJ1_YEAST	P53935 saccharomyc
45	69	26.7	338	1	LEGB_PEA	P14594 pisum sativ

ALIGNMENTS

```
RESULT   1
VCLB_GOSHI
ID   VCLB_GOSHI     STANDARD;      PRT;   588 AA.
AC   P09801;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT   "Developmental biochemistry of cottonseed embryogenesis and
RT   germination. XVIII. cDNA and amino acid sequences of the members of
RT   the storage protein families.";
RL   Plant Mol. Biol. 7:475-489(1986).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16891; AAA33071.1; -.
DR   PIR; A30838; FWCNAB.
DR   HSSP; P50477; 1CAX.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL        1     25
FT   CHAIN        26    588       VICILIN C72.
SQ   SEQUENCE   588 AA;  69729 MW;  63E699B29AB8ADEB CRC64;


  Query Match             48.1%;  Score 124;  DB 1;  Length 588;
  Best Local Similarity   46.3%;  Pred. No. 7.6e-05;
  Matches   19;  Conservative   12;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       4 DPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQ 44
           || ::|| |:|||| :   ::||:|||: |: :| |: :||
Db      36 DPPKRYEDCRRRCEWDTRGQKEQQQCEESCKSQYGEKDQQQ 76


RESULT   2
VCLA_GOSHI
ID   VCLA_GOSHI     STANDARD;      PRT;   605 AA.
```

```
AC   P09799;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT   "Developmental biochemistry of cottonseed embryogenesis and
RT   germination. XIX. Sequences and genomic organization of the alpha
RT   globulin (vicilin) genes of cottonseed.";
RL   Plant Mol. Biol. 9:533-546(1987).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M19378; AAA33069.1; -.
DR   PIR; S06398; S06398.
DR   HSSP; P50477; 1CAX.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL       1     23
FT   CHAIN       24    605       VICILIN GC72-A.
SQ   SEQUENCE   605 AA;  71049 MW;  C9DB9371C976953B CRC64;


  Query Match            48.1%;  Score 124;  DB 1;  Length 605;
  Best Local Similarity  43.5%;  Pred. No. 7.8e-05;
  Matches   20;  Conservative   13;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       2 ERDPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEE 47
           | ||:|:|| |::||: |   : ||::|| | | : ||:|::  |:
Db      34 EDDPQQRYEDCRKRCQLETRGQTEQDKCEDRSETQLKEEQQRDGED 79


RESULT   3
APA4_MACFA
ID   APA4_MACFA     STANDARD;      PRT;   429 AA.
AC   P33621;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV).
GN   APOA4.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEUKOCYTE;
RX   MEDLINE=93192330; PubMed=8448212;
RA   Osada J., Pocovi M., Nicolosi R.J., Schaefer E.J., Ordovas J.M.;
RT   "Nucleotide sequences of the Macaca fascicularis apolipoprotein C-III
RT   and A-IV genes.";
RL   Biochim. Biophys. Acta 1172:335-339(1993).
CC   -!- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC       CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC       LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC       COMPONENT OF HDL AND CHYLOMICRONS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SECRETED IN PLASMA.
CC   -!- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC       22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC       MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC       HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC       THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC       ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC   -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X68361; CAA48421.1; -.
DR   PIR; S29565; S29565.
DR   PIR; S30195; S30195.
DR   HSSP; P02649; 1NFO.
DR   INTERPRO; IPR000074; -.
DR   PFAM; PF01442; Apolipoprotein; 1.
KW   Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.
FT   SIGNAL       1     20       BY SIMILARITY.
FT   CHAIN       21    429       APOLIPOPROTEIN A-IV.
FT   DOMAIN      33    330       13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      33     54       1.
FT   REPEAT      60     81       2.
FT   REPEAT      82    103       3.
FT   REPEAT     115    136       4.
FT   REPEAT     137    158       5.
FT   REPEAT     159    180       6.
FT   REPEAT     181    202       7.
FT   REPEAT     203    224       8.
FT   REPEAT     225    246       9.
FT   REPEAT     247    268       10.
FT   REPEAT     269    286       11.
FT   REPEAT     287    308       12.
FT   REPEAT     309    330       13.
FT   DOMAIN     372    420       GLU/GLN-RICH.
SQ   SEQUENCE   429 AA;  49876 MW;  3D458F551D0DB60C CRC64;


  Query Match            33.7%;  Score 87;  DB 1;  Length 429;
  Best Local Similarity  34.8%;  Pred. No. 0.13;
  Matches   16;  Conservative   16;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       2 ERDPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEE 47
           |:   || || | : : :  :: :|:| ||: |:: :|||::|::|
Db     374 EQQREQQQEQQQEQEQEQQQQQEQQQQQEQQREQQQQEQQQEQQQE 419


RESULT   4
AGRP_LUFCY
ID   AGRP_LUFCY     STANDARD;      PRT;    47 AA.
AC   P56568;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   6.5 KDA ARGININE/GLUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).
OS   Luffa cylindrica (Smooth loofah) (Sponge gourd).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Cucurbitales; Cucurbitaceae; Luffa.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SEED;
RX   MEDLINE=97357433; PubMed=9214759;
RA   Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;
```

```
RT   "Primary structure of 6.5k-arginine/glutamate-rich polypeptide from
RT   the seeds of sponge gourd (Luffa cylindrica).";
RL   Biosci. Biotechnol. Biochem. 61:984-988(1997).
CC   -!- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
CC       RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CC   -!- MASS SPECTROMETRY: MW=5693.39; METHOD=MALDI.
CC   -!- SIMILARITY: SOME, TO 7S SEED STORAGE PROTEINS.
KW   Seed storage protein.
FT   DISULFID     12     33
FT   DISULFID     16     29
SQ   SEQUENCE   47 AA;  5698 MW;  588B0EC82273AC05 CRC64;


  Query Match             33.3%;  Score 86;  DB 1;  Length 47;
  Best Local Similarity   34.1%;  Pred. No. 0.024;
  Matches   14;  Conservative   11;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       5 PRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQE 45
           || :|| |: ||:        | :|:| ||:  :|:: ::|
Db       5 PRTEYEACRVRCQVAEHGVERQRRCQQVCEKRLREREGRRE 45


RESULT   5
TRHY_HUMAN
ID   TRHY_HUMAN     STANDARD;      PRT;  1898 AA.
AC   Q07283;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRICHOHYALIN.
GN   THH OR TRHY OR THL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93280194; PubMed=7685034;
RA   Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA   Steinert P.M.;
RT   "The structure of human trichohyalin. Potential multiple roles as a
RT   functional EF-hand-like calcium-binding protein, a cornified cell
RT   envelope precursor, and an intermediate filament-associated (cross-
RT   linking) protein.";
RL   J. Biol. Chem. 268:12164-12176(1993).
RN   [2]
RP   SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=93315897; PubMed=7686953;
RA   O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RT   "Trichohyalin: a structural protein of hair, tongue, nail, and
RT   epidermis.";
RL   J. Invest. Dermatol. 101:65S-71S(1993).
CC   -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC       IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC       INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC       LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC       ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC       WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC       ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC       ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC       DIFFERENTIATION.
CC   -!- SUBUNIT: MONOMER (PROBABLE).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC       THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC       THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC       THE EPIDERMIS.
CC   -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC       CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC       ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC       OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC       ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC       THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC       DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC       THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC       DIFFERENT SPECIES.
CC   -!- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC       PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09190; AAA65582.1; -.
DR   PIR; A45973; A45973.
DR   HSSP; P02633; 1BOC.
DR   MIM; 190370; -.
DR   INTERPRO; IPR001751; -.
DR   INTERPRO; IPR002048; -.
DR   PFAM; PF01023; S_100; 1.
DR   PFAM; PF00036; efhand; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00303; S100_CABP; 1.
KW   Repeat; Calcium-binding.
FT   DOMAIN       1     91       S-100 LIKE.
FT   CA_BIND     22     33       SITE I (LOW AFFINITY) (POTENTIAL).
FT   CA_BIND     62     73       SITE II (HIGH AFFINITY) (POTENTIAL).
FT   DOMAIN     314    390       6 X 13 AA TANDEM REPEATS OF
FT                               R-R-E-Q-E-E-E-R-R-E-Q-Q-L.
FT   REPEAT     314    326       1-1 (APPROXIMATE).
FT   REPEAT     327    339       1-2 (APPROXIMATE).
FT   REPEAT     340    351       1-3 (APPROXIMATE).
FT   REPEAT     352    364       1-4.
FT   REPEAT     365    377       1-5.
FT   REPEAT     378    390       1-6.
FT   DOMAIN     391    444       9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT   REPEAT     391    396       2-1.
FT   REPEAT     397    402       2-2.
FT   REPEAT     403    408       2-3.
FT   REPEAT     409    414       2-4.
FT   REPEAT     415    420       2-5.
FT   REPEAT     421    426       2-6.
FT   REPEAT     427    432       2-7.
FT   REPEAT     433    438       2-8.
FT   REPEAT     439    444       2-9.
FT   DOMAIN     444    702       9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT   DOMAIN     923   1162       8 X 30 AA TANDEM REPEATS.
FT   REPEAT     923    952       4-1.
FT   REPEAT     953    982       4-2.
FT   REPEAT     983   1012       4-3.
FT   REPEAT    1013   1042       4-4.
FT   REPEAT    1043   1072       4-5.
FT   REPEAT    1073   1102       4-6.
FT   REPEAT    1103   1132       4-7.
FT   REPEAT    1133   1162       4-8.
FT   DOMAIN    1250   1849       23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT   CONFLICT  1752   1752       F -> L (IN REF. 2).
FT   CONFLICT  1794   1801       QERDRQYR -> RSETGSTG (IN REF. 2).
FT   CONFLICT  1857   1857       Q -> K (IN REF. 2).
FT   CONFLICT  1880   1880       V -> G (IN REF. 2).
SQ   SEQUENCE   1898 AA;  247219 MW;  A74B5947FB62E31D CRC64;


  Query Match             33.3%;  Score 86;  DB 1;  Length 1898;
  Best Local Similarity   40.9%;  Pred. No. 0.59;
  Matches   18;  Conservative    9;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       2 ERDPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQE 45
           ||   |   : | | |     | |:|: ||| :||::|::|:||
Db     660 ERRHEQLRREQQERREQRLKREEEEERLEQRLKREHEEERREQE 703
```

```
RESULT   6
TRHY_RABIT
ID   TRHY_RABIT     STANDARD;      PRT;  1407 AA.
AC   P37709;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRICHOHYALIN.
GN   THH.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fietz M.J., Rogers G.E.;
RL   Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC       IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC       INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC       LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC       ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC       WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC       ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC       ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC       DIFFERENTIATION.
CC   -!- SUBUNIT: HOMODIMER (PROBABLE).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC       THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC       THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC       THE EPIDERMIS.
CC   -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC       CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC       ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC       OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC       ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC       THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC       DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC       THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC       DIFFERENT SPECIES.
CC   -!- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC       PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z19092; CAA79519.1; -.
DR   PIR; S28589; S28589.
DR   HSSP; P02633; 1BOC.
DR   INTERPRO; IPR001751; -.
DR   INTERPRO; IPR002048; -.
DR   PFAM; PF01023; S_100; 1.
DR   PFAM; PF00036; efhand; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00303; S100_CABP; 1.
KW   Repeat; Calcium-binding.
FT   DOMAIN        1     91       S-100 LIKE.
FT   CA_BIND      22     33       SITE I (LOW AFFINITY) (POTENTIAL).
FT   CA_BIND      62     73       SITE II (HIGH AFFINITY) (POTENTIAL).
SQ   SEQUENCE   1407 AA;  183781 MW;  AE17D2A159F12B7F CRC64;


  Query Match             32.8%;  Score 84.5;  DB 1;  Length 1407;
  Best Local Similarity   44.4%;  Pred. No. 0.62;
  Matches   24;  Conservative   10;  Mismatches   11;  Indels    9;  Gaps    4;

Qy       2 ERDPR---QQY--EQCQRRCESEATEEREQEQCEQRCEREYKEQQ--RQQEEE 47
           ||| |    :|:  |:| || | |  :| || :  :| ||:::|:|  ||||||
Db    1304 ERDRRYRAEEQFAREEKSRRLERELRQEEEQRRRRER-ERKFREEQLRRQQEEE 1356


RESULT   7
APA4_PAPAN
ID   APA4_PAPAN     STANDARD;      PRT;   401 AA.
AC   Q28758;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV) (FRAGMENT).
GN   APOA4.
OS   Papio anubis (Olive baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RX   MEDLINE=93340170; PubMed=8101842;
RA   Hixson J.E., Kammerer C.M., Mott G.E., Britten M.L., Birnbaum S.,
RA   Powers P.K., Vandeberg J.L.;
RT   "Baboon apolipoprotein A-IV. Identification of Lys-76-->Glu that
RT   distinguishes two common isoforms and detection of length
RT   polymorphisms at the carboxyl terminus.";
RL   J. Biol. Chem. 268:15667-15673(1993).
CC   -!- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC       CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC       LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC       COMPONENT OF HDL AND CHYLOMICRONS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED PRIMARILY IN THE INTESTINE.
CC       SECRETED IN PLASMA.
CC   -!- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC       22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC       MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC       HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC       THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC       ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC   -!- POLYMORPHISM: THERE ARE TWO COMMON APOA-IV ISOFORMS, I (SHOWN
CC       HERE) AND E. THE I ISOFORM IS ASSOCIATED WITH HIGHER LEVELS OF
CC       HIGH DENSITY LIPOPROTEIN-C ON A HIGH CHOLESTEROL, SATURATED FAT
CC       DIET.
CC   -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13174; AAA35379.1; -.
DR   HSSP; P02649; 1NFO.
DR   INTERPRO; IPR000074; -.
DR   PFAM; PF01442; Apolipoprotein; 1.
KW   Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal;
KW   Polymorphism.
FT   NON_TER       1      1
FT   SIGNAL       <1      4       POTENTIAL.
FT   CHAIN         5    401       APOLIPOPROTEIN A-IV.
FT   DOMAIN       17    314       13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT       17     38       1.
FT   REPEAT       44     65       2.
FT   REPEAT       66     87       3.
FT   REPEAT       99    120       4.
FT   REPEAT      121    142       5.
```

```
FT   REPEAT      143    164       6.
FT   REPEAT      165    186       7.
FT   REPEAT      187    208       8.
FT   REPEAT      209    230       9.
FT   REPEAT      231    252       10.
FT   REPEAT      253    270       11.
FT   REPEAT      271    292       12.
FT   REPEAT      293    314       13.
FT   DOMAIN      356    394       GLU/GLN-RICH.
FT   VARIANT      80     80       K -> E (IN ISOFORM E).
SQ   SEQUENCE   401 AA;  46538 MW;  0A76D1284AA9837F CRC64;


  Query Match             30.8%;  Score 79.5;  DB 1;  Length 401;
  Best Local Similarity   36.4%;  Pred. No. 0.59;
  Matches   16;  Conservative   13;  Mismatches    8;  Indels    7;  Gaps    1;

Qy      4 DPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEE 47
          :| || || |        :|::||| |:: ::| :::| ||:|:
Db    356 EPEQQQEQQQE-------QEQQQEQQEEQQQQEQQQEQEQQQEQ 392


RESULT   8
ABRA_PLAFC
ID   ABRA_PLAFC     STANDARD;      PRT;   743 AA.
AC   P22620;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   101 KDA MALARIA ANTIGEN (P101) (ACIDIC BASIC REPEAT ANTIGEN).
GN   ABRA.
OS   Plasmodium falciparum (isolate Camp / Malaysia).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88298794; PubMed=3042768;
RA   Weber J.L., Lyon J.A., Wolff R.H., Hall T., Lowell G.H., Chulay J.D.;
RT   "Primary structure of a Plasmodium falciparum malaria antigen located
RT   at the merozoite surface and within the parasitophorous vacuole.";
RL   J. Biol. Chem. 263:11421-11425(1988).
CC   -!- SUBCELLULAR LOCATION: AT THE MEROZOITE SURFACE AND WITHIN THE
CC       PARASITOPHOROUS VACUOLE.
CC   -!- PTM: NOT GLYCOSYLATED (PROBABLE).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J03902; AAA29462.1; -.
DR   PIR; A29232; A29232.
KW   Antigen; Malaria; Repeat.
FT   DOMAIN      226    273       8 X 6 AA REPEATS OF [VT]-N-D-[ED]-[ED]-D.
FT   REPEAT      226    231
FT   REPEAT      232    237
FT   REPEAT      238    243
FT   REPEAT      244    249
FT   REPEAT      250    255
FT   REPEAT      256    261
FT   REPEAT      262    267
FT   REPEAT      268    273
FT   DOMAIN      674    731       TANDEM REPEATS OF K-E(2,3).
SQ   SEQUENCE   743 AA;  86622 MW;  FBF0B0F8B07D922C CRC64;


  Query Match             30.6%;  Score 79;  DB 1;  Length 743;
  Best Local Similarity   30.4%;  Pred. No. 1.1;
  Matches   14;  Conservative   19;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 ERDPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEE 47
            |::  :: |: ::  | :  ||:|:|: |:  |:| ||:::::||
Db    678 EKEKEKEKEKEEKEKEEKEKEEKEKEKEEKEKEKEEKEEEKKEKEE 723


RESULT   9
TAGB_DICDI
ID   TAGB_DICDI     STANDARD;      PRT;  1905 AA.
AC   P54683;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21.-).
GN   TAGB.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Dictyosteliida; Dictyostelium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=95262903; PubMed=7744252;
RA   Shaulsky G., Kuspa A., Loomis W.F.;
RT   "A multidrug resistance transporter/serine protease gene is required
RT   for prestalk specialization in Dictyostelium.";
RL   Genes Dev. 9:1111-1122(1995).
CC   -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
CC       INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC       S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC       TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   -!- SIMILARITY: STRONG, TO TAGC.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U20432; AAA62212.1; -.
DR   HSSP; P13569; 1NBD.
DR   DICTYDB; DD02059; TAGB.
DR   INTERPRO; IPR000209; -.
DR   INTERPRO; IPR001140; -.
DR   INTERPRO; IPR001617; -.
DR   PFAM; PF00664; ABC_membrane; 1.
DR   PFAM; PF00005; ABC_tran; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW   Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?   1905       PRESTALK-SPECIFIC PROTEIN TAGB.
FT   DOMAIN      378    700       PROTEASE.
FT   DOMAIN        ?   1905       ABC TRANSPORTER.
FT   TRANSMEM   1011   1031       POTENTIAL.
FT   TRANSMEM   1076   1096       POTENTIAL.
FT   TRANSMEM   1121   1141       POTENTIAL.
FT   TRANSMEM   1210   1230       POTENTIAL.
FT   TRANSMEM   1309   1329       POTENTIAL.
FT   TRANSMEM   1332   1352       POTENTIAL.
FT   ACT_SITE    387    387       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    432    432       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    695    695       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1553   1560       ATP (POTENTIAL).
FT   DOMAIN       63     67       POLY-GLN.
FT   DOMAIN       95    104       POLY-ASN.
FT   DOMAIN      107    134       POLY-ASN.
FT   DOMAIN      311    321       POLY-SER.
FT   DOMAIN      833    839       POLY-SER.
```

```
FT   DOMAIN      838    844       POLY-GLY.
FT   DOMAIN      871    876       POLY-LEU.
FT   DOMAIN     1012   1015       POLY-ILE.
FT   DOMAIN     1386   1389       POLY-GLU.
FT   DOMAIN     1398   1404       POLY-GLY.
FT   DOMAIN     1445   1450       POLY-ASN.
FT   DOMAIN     1765   1779       POLY-ASN.
FT   DOMAIN     1782   1785       POLY-SER.
FT   DOMAIN     1807   1812       POLY-PRO.
FT   DOMAIN     1815   1860       POLY-GLN.
FT   DOMAIN     1872   1878       POLY-PRO.
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    747    747       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    823    823       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1172   1172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1522   1522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1658   1658       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1905 AA;  212518 MW;  B8E223FA8B9AE13C CRC64;


  Query Match             30.6%;  Score 79;  DB 1;  Length 1905;
  Best Local Similarity   25.6%;  Pred. No. 2.5;
  Matches   11;  Conservative   23;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       5 PRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEE 47
           |::| || ::: : :  ::::|:| :|: ::: ::||:||:::
Db    1812 PQEQQEQQEQQQQQQQEQQQQQQQQQQQQQQQQQQQQQQQQQ 1854


RESULT  10
IF2_SALTY
ID   IF2_SALTY      STANDARD;      PRT;   892 AA.
AC   Q9ZF31;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRANSLATION INITIATION FACTOR IF-2.
GN   INFB.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RA   Fage-Larsen J., Steffensen S.A.D.A., Hedegaard J., Olsen J.E.,
RA   Mortensen K.K., Sperling-Petersen H.U.;
RT   "Sequence of the infB gene from Salmonella typhimurium.";
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
CC       OF PROTEIN SYNTHESIS IN VITRO, PROTECTS FORMYLMETHIONYL-TRNA FROM
CC       SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S
CC       RIBOSOMAL SUBUNITS. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
CC       DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- ALTERNATIVE PRODUCTS: USING ALTERNATIVE INITIATION CODONS IN THE
CC       SAME READING FRAME, THE GENE TRANSLATES INTO THREE ISOZYMES:
CC       ALPHA, BETA AND GAMMA.
CC   -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ002552; CAA05549.1; -.
DR   EMBL; AJ002552; CAA05550.1; -.
DR   EMBL; AJ002552; CAA05551.1; -.
DR   STYGENE; SG10757; INFB.
DR   INTERPRO; IPR000178; -.
DR   INTERPRO; IPR000795; -.
DR   PFAM; PF00009; GTP_EFTU; 1.
DR   PFAM; PF02131; IF2; 1.
DR   PROSITE; PS01176; IF2; 1.
KW   Initiation factor; Protein biosynthesis; GTP-binding;
KW   Alternative initiation.
FT   CHAIN         1    892       INITIATION FACTOR IF-2-ALPHA.
FT   CHAIN       159    892       INITIATION FACTOR IF-2-BETA.
FT   CHAIN       167    892       INITIATION FACTOR IF-2-GAMMA.
FT   INIT_MET    159    159       FOR IF2-BETA.
FT   INIT_MET    167    167       FOR IF2-GAMMA.
FT   DOMAIN      394    542       G-DOMAIN.
FT   NP_BIND     400    407       GTP (BY SIMILARITY).
FT   NP_BIND     446    450       GTP (BY SIMILARITY).
FT   NP_BIND     500    503       GTP (BY SIMILARITY).
SQ   SEQUENCE   892 AA;  97388 MW;  9472B20DD55597B1 CRC64;


  Query Match             30.2%;  Score 78;  DB 1;  Length 892;
  Best Local Similarity   33.3%;  Pred. No. 1.6;
  Matches   18;  Conservative   12;  Mismatches   16;  Indels    8;  Gaps    2;

Qy       2 ERDPRQ-----QYEQCQRRCESEATEEREQE---QCEQRCEREYKEQQRQQEEE 47
           :|||::       || ||  | :|  | |::   : :|: |||  || :::  |
Db      94 KRDPQEAERLAAEEQAQREAEEQARREAEEQAKREAQQKAEREAAEQAKREAAE 147


RESULT  11
SNF5_YEAST
ID   SNF5_YEAST      STANDARD;      PRT;   905 AA.
AC   P18480;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRANSCRIPTION REGULATORY PROTEIN SNF5 (SWI/SNF COMPLEX COMPONENT SNF5)
DE   (TRANSCRIPTION FACTOR TYE4).
GN   SNF5 OR TYE4 OR SWI10 OR YBR289W OR YBR2036.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MCY;
RX   MEDLINE=91042489; PubMed=2233708;
RA   Laurent B.C., Treitel M.A., Carlson M.;
RT   "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and
RT   proline-rich transcriptional activator that affects expression of a
RT   broad spectrum of genes.";
RL   Mol. Cell. Biol. 10:5616-5625(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94378722; PubMed=8091861;
RA   Holmstroem K., Brandt T., Kallesoe T.;
RT   "The sequence of a 32,420 bp segment located on the right arm of
RT   chromosome II from Saccharomyces cerevisiae.";
RL   Yeast 10:S47-S62(1994).
CC   -!- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
CC       COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
CC       OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
CC       BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
CC   -!- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE SNF5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M36482; AAA35062.1; -.
DR   EMBL; X76053; CAA53652.1; -.
DR   EMBL; Z36158; CAA85254.1; -.
DR   PIR; S44551; RGBYS5.
DR   PIR; S39145; S39145.
DR   SGD; S0000493; SNF5.
KW   Transcription regulation; Activator; Nuclear protein.
FT   DOMAIN       31    270       GLN-RICH.
FT   DOMAIN       72    132       PRO-RICH.
FT   DOMAIN      272    324       PRO-RICH.
FT   DOMAIN      489    588       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      714    882       PRO-RICH.
FT   DOMAIN      755    798       ARG/LYS-RICH (BASIC).
FT   CONFLICT    564    564       E -> D (IN REF. 1).
SQ   SEQUENCE   905 AA;  102557 MW;  A287B4A648DD1A35 CRC64;


  Query Match             30.2%;  Score 78;  DB 1;  Length 905;
  Best Local Similarity   24.4%;  Pred. No. 1.6;
  Matches   11;  Conservative   24;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 ERDPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEE 46
           ::  :||::| |:: : :  ::::|:| :|: ::: ::||:||::
Db     224 QQQQQQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 268


RESULT  12
PRO_DROME
ID   PRO_DROME      STANDARD;      PRT;  1403 AA.
AC   P29617;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROTEIN PROSPERO.
GN   PROS.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92069760; PubMed=1720353;
RA   Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
RT   "Prospero is expressed in neuronal precursors and encodes a nuclear
RT   protein that is involved in the control of axonal outgrowth in
RT   Drosophila.";
RL   Cell 67:941-953(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92171948; PubMed=1540176;
RA   Matsuzaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;
RT   "Cloning of the Drosophila prospero gene and its expression in
RT   ganglion mother cells.";
RL   Biochem. Biophys. Res. Commun. 182:1326-1332(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93083413; PubMed=1842358;
RA   Chu-Lagraff Q., Wright D.M., McNeil L.K., Doe C.Q.;
RT   "The prospero gene encodes a divergent homeodomain protein that
RT   controls neuronal identity in Drosophila.";
RL   Development Suppl. 2:79-85(1991).
RN   [4]
RP   SIMILARITY TO C.ELEGANS CEH-26.
RX   MEDLINE=94212446; PubMed=7909177;
RA   Buerglin T.R.;
RT   "A Caenorhabditis elegans prospero homologue defines a novel domain.";
RL   Trends Biochem. Sci. 19:70-71(1994).
CC   -!- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR
CC       GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS
CC       CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED
CC       FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR
CC       PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT
CC       FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE
CC       TRANSCRIPTION BY BINDING TO DNA.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PROS-L (SHOWN HERE) AND PROS-
CC       S; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT N-
CC       TERMINAL AMINO ACIDS OF THE HOMEODOMAIN.
CC   -!- TISSUE SPECIFICITY: NEURONAL PRECURSORS. EXPRESSED IN THE
CC       DEVELOPING CNS, LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING
CC       FORMATION.
CC   -!- SIMILARITY: BELONGS TO THE PROSPERO FAMILY OF HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81389; AAA28841.1; -.
DR   EMBL; D10609; BAA01464.1; -.
DR   EMBL; Z11743; CAA77802.1; -.
DR   PIR; A41089; A41089.
DR   PIR; JQ1397; JQ1397.
DR   FLYBASE; FBgn0004595; pros.
KW   Nuclear protein; Transcription regulation; DNA-binding; Homeobox;
KW   Developmental protein; Alternative splicing.
FT   DOMAIN        4     12       POLY-ALA.
FT   DOMAIN       28     31       POLY-SER.
FT   DOMAIN       32     35       POLY-ASN.
FT   DOMAIN      188    191       POLY-ALA.
FT   DOMAIN      253    260       POLY-GLN.
FT   DOMAIN      270    276       POLY-ASN.
FT   DOMAIN      282    286       POLY-ASN.
FT   DOMAIN      431    437       POLY-ASP.
FT   DOMAIN      505    508       POLY-ALA.
FT   DOMAIN      717    737       POLY-GLN.
FT   DOMAIN      754    763       POLY-GLN.
FT   DOMAIN      766    772       POLY-GLN.
FT   DOMAIN      934    937       POLY-ALA.
FT   DOMAIN      952    957       POLY-GLN.
FT   DOMAIN      960    963       POLY-GLN.
FT   DOMAIN      966    970       POLY-GLN.
FT   DOMAIN      991    998       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN     1029   1048       POLY-GLN.
FT   DOMAIN     1074   1082       POLY-PRO.
FT   DNA_BIND   1241   1303       HOMEOBOX (ATYPICAL).
FT   DOMAIN     1304   1403       PROSPERO-LIKE.
FT   VARSPLIC   1216   1244       MISSING (IN ISOFORM PROS-S).
FT   CONFLICT     76     98       AKMLNELFGRQMKQAQDATSGLP -> GQDAERAVWPPDEA
FT                                GPGRNEWPA (IN REF. 1).
FT   CONFLICT    120    144       IGSLNSTSKLLQQQHNNNSIAPANS -> NLALQFHVQVAA
FT                                AAAITTALLPPIG (IN REF. 1).
FT   CONFLICT    418    418       H -> Q (IN REF. 2).
FT   CONFLICT    677    677       A -> C (IN REF. 1).
FT   CONFLICT    802    802       A -> S (IN REF. 1).
FT   CONFLICT    958    958       T -> S (IN REF. 1).
FT   CONFLICT   1048   1048       Q -> QQQQQ (IN REF. 1).
SQ   SEQUENCE   1403 AA;  153569 MW;  9EFB9973E24E238E CRC64;


  Query Match             30.2%;  Score 78;  DB 1;  Length 1403;
  Best Local Similarity   29.4%;  Pred. No. 2.4;
  Matches   15;  Conservative   19;  Mismatches   13;  Indels    4;  Gaps    1;

Qy       1 YERDPRQQYEQCQRRCESEATEEREQEQCEQ----RCEREYKEQQRQQEEE 47
           :|:: |   |  ::: : :  ::::|:| :|    | |:| :||||::||:
Db     704 FEQEARMAKEAAEQQQQQQQQQQQQQQQQQQQEQQRRFEQEQQEQQRRKEEQ 754


RESULT  13
```

```
SKS1_YEAST
ID   SKS1_YEAST     STANDARD;      PRT;   502 AA.
AC   Q12505;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SERINE/THREONINE-PROTEIN KINASE SKS1 (EC 2.7.1.-) (SUPPRESSOR KINASE
DE   OF SNF3).
GN   SKS1 OR SHA3 OR YPL026C OR LPB5.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97103770; PubMed=8948096;
RA   Yang Z., Bisson L.F.;
RT   "The SKS1 protein kinase is a multicopy suppressor of the snf3
RT   mutation of Saccharomyces cerevisiae.";
RL   Yeast 12:1407-1419(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wang Y., Ahmed A., Bussey H., Fortin N., Friesen J.D., Hall J.,
RA   Storms R.K., Vo D.H., Winnett E.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY HAVE A ROLE IN GLUCOSE REGULATION.
CC   -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC       PROTEIN KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U30613; AAC49570.1; -.
DR   EMBL; U36624; AAB68161.1; -.
DR   HSSP; P24941; 1AQ1.
DR   SGD; S0005947; SKS1.
DR   INTERPRO; IPR000719; -.
DR   INTERPRO; IPR002290; -.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN       10    338       PROTEIN KINASE.
FT   NP_BIND      16     24       ATP (BY SIMILARITY).
FT   BINDING      39     39       ATP.
FT   ACT_SITE    186    186       BY SIMILARITY.
FT   MUTAGEN      39     39       K->R: LOSS OF ACTIVITY.
SQ   SEQUENCE   502 AA;  57844 MW;  B106D084BAFA61E5 CRC64;


  Query Match             30.0%;  Score 77.5;  DB 1;  Length 502;
  Best Local Similarity   31.9%;  Pred. No. 1.1;
  Matches   15;  Conservative   15;  Mismatches   14;  Indels    3;  Gaps    1;

Qy       1 YERDPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEE 47
           :  |  || :| |   : :  :|:|||| ::: : : : ||:|:||:
Db     372 FSADQEQQQQQQQ---QQQQVQEQEQEQKQEQIQNQEQAQQQQEEED 415


RESULT  14
TF2D_HUMAN
ID   TF2D_HUMAN     STANDARD;      PRT;   339 AA.
AC   P20226;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR) (TATA
DE   SEQUENCE-BINDING PROTEIN) (TBP).
GN   TBP OR TFIID OR TF2D.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90302010; PubMed=2194289;
RA   Kao C.C., Lieberman P.M., Schmidt M.C., Zhou Q., Pei R., Berk A.J.;
RT   "Cloning of a transcriptionally active human TATA binding factor.";
RL   Science 248:1646-1649(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90326195; PubMed=2374612;
RA   Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,
RA   Roeder R.G.;
RT   "Highly conserved core domain and unique N terminus with presumptive
RT   regulatory motifs in a human TATA factor (TFIID).";
RL   Nature 346:387-390(1990).
RN   [3]
RP   DOMAINS.
RX   MEDLINE=90302006; PubMed=2363050;
RA   Peterson M.G., Tanese N., Pugh B.F., Tjian R.;
RT   "Functional domains and upstream activation properties of cloned
RT   human TATA binding protein.";
RL   Science 248:1625-1630(1990).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337.
RX   MEDLINE=96209823; PubMed=8643494;
RA   Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,
RA   Burley S.K.;
RT   "Crystal structure of a human TATA box-binding protein/TATA element
RT   complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4862-4867(1996).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339.
RX   MEDLINE=96346176; PubMed=8757291;
RA   Juo Z.S., Chiu T.K., Leiberman P.M., Baikalov I., Berk A.J.,
RA   Dickerson R.E.;
RT   "How proteins recognize the TATA box.";
RL   J. Mol. Biol. 261:239-254(1996).
CC   -!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC       OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIID BINDS
CC       SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO
CC       THE POSITION OF TRANSCRIPTION INITIATION.
CC   -!- SUBUNIT: BINDS DNA AS A MONOMER.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: THE C-TERMINAL 180 RESIDUES ARE EXTREMELY WELL
CC       CONSERVED IN ALL EUKARYOTIC TFIID.
CC   -!- SIMILARITY: WEAK, WITH BACTERIAL POLYMERASE SIGMA-FACTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M55654; AAA36731.1; -.
DR   EMBL; X54993; CAA38736.1; -.
DR   PIR; A34830; A34830.
DR   PIR; A34831; A34831.
DR   PDB; 1TGH; 01-AUG-96.
DR   PDB; 1CDW; 23-DEC-96.
DR   TRANSFAC; T00794; -.
DR   MIM; 600075; -.
DR   INTERPRO; IPR000814; -.
DR   PFAM; PF00352; TBP; 2.
DR   PRINTS; PR00686; TIFACTORIID.
DR   PROSITE; PS00351; TFIID; 2.
KW   Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW   3D-structure.
FT   DOMAIN       55     95       POLY-GLN.
```

```
FT   REPEAT      165    241       1.
FT   REPEAT      255    332       2.
FT   CONFLICT     58     61       MISSING (IN REF. 2).
SQ   SEQUENCE   339 AA;  37698 MW;  A61A578D972B970B CRC64;


  Query Match             29.8%;  Score 77;  DB 1;  Length 339;
  Best Local Similarity   30.2%;  Pred. No. 0.86;
  Matches   13;  Conservative   18;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      2 ERDPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQ 44
          |   ||| :| |:: : :  ::::|:| :|: ::: ::||:|| 
Db     53 EEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 95


RESULT  15
NFM_CHICK
ID   NFM_CHICK      STANDARD;      PRT;   857 AA.
AC   P16053;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)
DE   (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M).
GN   NEFM.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90174973; PubMed=2106668;
RA   Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
RT   "Isolation of the chicken middle-molecular weight neurofilament
RT   (NF-M) gene and characterization of its promoter.";
RL   Nucleic Acids Res. 18:521-529(1990).
RN   [2]
RP   SEQUENCE OF 259-857 FROM N.A.
RX   MEDLINE=88112814; PubMed=3123320;
RA   Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT   "Identification of gene products expressed in the developing chick
RT   visual system: characterization of a middle-molecular-weight
RT   neurofilament cDNA.";
RL   Genes Dev. 1:699-708(1987).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17102; CAA34958.1; -.
DR   EMBL; X05558; CAA29073.1; -.
DR   PIR; A27040; A27040.
DR   PIR; S08061; S08061.
DR   PIR; S15762; S15762.
DR   INTERPRO; IPR001664; -.
DR   PFAM; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW   Phosphorylation; Glycoprotein.
FT   INIT_MET      0      0
FT   DOMAIN        1     98       HEAD.
FT   DOMAIN       99    406       ROD.
FT   DOMAIN      407    857       TAIL.
FT   DOMAIN       99    130       COIL 1A.
FT   DOMAIN      131    143       LINKER 1.
FT   DOMAIN      144    242       COIL 1B.
FT   DOMAIN      243    259       LINKER 12.
FT   DOMAIN      260    281       COIL 2A.
FT   DOMAIN      282    285       LINKER 2.
FT   DOMAIN      286    406       COIL 2B.
FT   CARBOHYD     46     46       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   CARBOHYD    426    426       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   CONFLICT    546    546       G -> R (IN REF. 2).
SQ   SEQUENCE   857 AA;  95704 MW;  4E2E0FC6AC64778B CRC64;


  Query Match             29.8%;  Score 77;  DB 1;  Length 857;
  Best Local Similarity   39.5%;  Pred. No. 1.9;
  Matches   17;  Conservative    8;  Mismatches   16;  Indels    2;  Gaps    1;

Qy      7 QQYEQCQRRCESEATEER--EQEQCEQRCEREYKEQQRQQEEE 47
          |: || : : | || ||    ::  ||  | | ||::  :|||
Db    481 QEEEQEEEKAEEEAVEEEAVSEKAAEQAAEEEEKEEEEAEEE 523


Search completed: March  1, 2001, 16:16:52
Job time: 432 sec
```